Parameter estimation for knowledge and diagnosis of electrical machines

Jean-Claude Trigeassou, Thierry Poinot, Smail Bachir

Chapitre 1

Parameter estimation for knowledge and diagnosis of electrical machines

Jean-Claude Trigeassou[1], **Thierry Poinot**[1] **and Smail Bachir**[2]

[1]University of Poitiers, LAII Laboratory
Avenue du Recteur Pineau
86022 Poitiers cedex, France
Emails : jean.claude.trigeassou@univ-poitiers.fr, thierry.poinot@univ-poitiers.fr

[2]University of Poitiers, XLIM Laboratory, SIC department
Bât. SP2MI, Téléport 2, Bvd Marie et Pierre Curie, 86962 Futuroscope Chasseneuil Cedex, France
URL : http ://www.sic.sp2mi.univ-poitiers.fr/
Email : smail.bachir@univ-poitiers.fr

.

Chapter 1

Parameter estimation for knowledge and diagnosis of electrical machines

1.1. Introduction

In automatic control, modeling and identification are fundamental and essential disciplines, which precede all operations of simulation, observation, synthesis of a control law or system monitoring. This stage of modeling and identification depends highly on the system and on the considered application. Thus, concerning synthesis of a control law, a discrete model associated with an algorithm derived from the least squares technique can be suitable. In an objective of simulation or state-observation, the same methodology can still be satisfactory. On the other hand, if the user wants to simulate the dynamic behavior of the system and to simulate the influence of some characteristic parameters at the same time, the approach using discrete model is insufficient and it will be necessary to use a continuous-time representation with differential equations nearer to the physical nature of the system. Finally, in an objective of monitoring, it is sure that a continuous-time model is preferable, especially when the user wishes to carry out a diagnosis of the system state starting from parameter estimation representative of its safe or fault functioning.

In case of discrete models where the parameters and the structure can have lost links with physics, one commonly speaks about black-box models. On the other hand, when the model is of continuous type governed by differential equations and when the parameters and the structure can approach physical equations, one speaks in this case

Chapter written by Jean-Claude TRIGEASSOU and Thierry POINOT and Smail BACHIR.

about knowledge models or more reasonably about grey-box models. Indeed, it is unrealistic to describe the physical behaviours exhaustively and in fact only approximate models (or reduced order models) are used, hence the term of grey-box model which better represents the modesty of the ambitions of the researcher.

In electrical engineering, even for a control application, the user prefers to use a model close to the reality, whose parameters (resistances, inductances...) have a tangible signification. One finds the same attitude in process engineering or in chemical engineering (for the constant kinetics of chemical reaction for example) like in robotics (for parameters such as masses and stiffness).

In this chapter, we are interested by the parameter estimation of electrical machines starting from continuous-time models, close to physics. The considered applications concern the estimation of electrical parameters, and particularly the monitoring of electrical machines based on parameter estimation. However, the same identification algorithms could be used for the state observation or the synthesis of control laws.

In this context, special attention will be given to system identification using continuous-time models. Depending on the nature of the residuals, two types of algorithms are commonly used: Equation-Error and Output-Error. In practice, Equation-Error algorithms are suitable only for models of the differential equation type having constant coefficients. For such models, many techniques [MEN 99] were imagined in order to express linearly the system output with respect to its parameters (L.P. model). Then, this property of linearity enables the use of least-squares methods whose essential interest is to provide an analytical expression of the estimated parameters [LJU 87]. Unfortunately, it is shown that for any L.P. model whose regressor depends directly (or indirectly by filtering) on the output values, the residuals are of the Equation-Error type and consequently the estimator is biased. A solution to eliminate this bias consists in using an instrumental variable technique [YOU 70, SöD 83]. This procedure complicates the identification algorithm and convergence can, in some cases, present some problems. Finally, because in general electrical machines are not governed by differential equations with constant coefficients but rather by nonlinear differential systems, it can be deduced that this identification methodology is not really adapted for the considered problem. However, these methods should not be rejected. Indeed, although their estimates are open to criticism, they can be used to initialize the Output-Error methods. The reader interested by this subject will be able to refer to some synthetic presentations [JEM 97, TRI 01] and in particular with a comparative study of these identification methods [MEN 99] where selection criteria are presented.

Thus, this chapter is dedicated to the presentation of the second category of algorithms which are of Output-Error type; in France they are also called "méthode du modèle" (model method), according to a term imposed by their promoter J. Richalet [RIC 71]. These algorithms present a wide range of applications; since they do not stay on restrictive linearity assumptions, they can be used for nonlinear differential

systems. They are not only used in electrical engineering but also in process and chemical engineering. Unfortunately, it is important to specify that they require a computational load much higher than the algorithms derived from least-squares. This explains certainly the interest carried by many researchers with all the alternatives of least-squares evoked previously. It also explains why, although the theory of these algorithms was proposed in the Sixties, it could truly start to impose itself only recently thanks to the rise in computing power of digital calculators.

The computational load of Output-Error algorithms is mainly due to the iterative nature of the minimization of a quadratic criterion by a Nonlinear Programming algorithm (N.L.P.) [HIM 72]. Indeed, at each iteration, it is necessary to simulate the model of the system together with its sensitivity model, according to the considered variant. Moreover, the convergence to a global optimum is not guaranteed, because of the non convexity of the quadratic criterion. Beyond these difficulties, which can be fortunately overcome, the principal interest of Output-Error algorithms is to provide an unbiased estimator, and (as opposed to Equation-Error techniques) with a certain immunity with respect to the modeling errors. However, it is necessary to examine the influence of this modeling error with respect to a deterministic parametric bias. In the same way, these Output-Error algorithms must be modified, if necessary, when the system is in closed-loop, inherent for example with the functioning of modern electrical machines.

An important problem is the problem of *a priori* knowledge on the parameters. This knowledge is necessary for the initialization of the N.L.P. algorithm, but in some cases, we can arrive at the paradox that the obtained estimator proves to be nonsensical when compared to the initial physical knowledge. Thus, we propose a methodology that allows the introduction of *a priori* knowledge in the estimator, according to the general principle of the Bayesian estimation [PET 81].

Although this last methodology must be used with understanding as it will be specified, it can be useful to improve convergence of the algorithm each time that an initial and reliable knowledge is available, particularly within the framework of machines monitoring, using an extended model composed of the safe functioning model of the system (thus with a good *a priori* knowledge) and of a model dedicated to a type of fault.

This chapter is composed of four sections. The first section gives the general methodology of Output-Error identification. The second section presents an approach that allows introduction of *a priori* knowledge in practice within the framework of the Bayesian estimation. An application to the parameter estimation of an induced machine in the Park's reference is given in the third section. Finally, a monitoring methodology of an electrical system with fault model and *a priori* knowledge of the safe functioning are the subject of the last section. This monitoring methodology

based on parameter estimation is used and developed in chapter 8 dedicated to the monitoring of electrical machine.

1.2. Identification using Output-Error algorithms

1.2.1. *Introduction*

These algorithms are known in France as "méthode du modèle" [RIC 71, RIC 91]. They are fundamentally characterized by the simulation of the model output based on the only knowledge of the input (Error-Equation algorithms are based on a prediction of the output using the knowledge of the input and past values of the output). Using this procedure, the simulated output is independent of the perturbation affecting the system (if the system is in open-loop); then, residuals correspond to this perturbation, hence the term of Output-Error and some interesting properties of convergence. On the other hand, this simulation complicates the problem of minimization of the criterion which requires the use of non-linear optimization techniques. Many approaches can be used with Output-Error algorithms. After some brief recall on the properties of Least-Squares algorithm used in Output-Error, we present the case where the gradient is computed by using the sensitivity functions.

1.2.2. *Least-Squares algorithm in Output-Error*

In some particular cases, the system output $y(t)$ is linear with respect to its parameters. Let this output model be $y = f(\underline{\theta}, u)$, where $u(t)$ is the input and $\underline{\theta}$ the parameter vector. The model is linear with respect to its parameters (L.P.) [WAL 97] if $y(t)$ can be written as follows:

$$y = \underline{\varphi}^T(u)\,\underline{\theta} \tag{1.1}$$

Where $\underline{\varphi}(u)$ is the regressor vector.

Let $\widehat{\underline{\theta}}$ be an estimation of $\underline{\theta}$. Then, using $u(t)$ (or u_k known at each sampling time t_k), the estimated output $\widehat{y}_k$ is obtained:

$$\widehat{y}_k = \underline{\varphi}_k^T(u)\,\widehat{\underline{\theta}} \tag{1.2}$$

Consider that we have K experimental data pairs $\{u_k, y_k^*\}$, obtained with sampling period T_e, such that $t = kT_e$. A quadratic criterion can thus be defined by:

$$J = \sum_{k=1}^{K} \varepsilon_k^2 = \sum_{k=1}^{K} \left(y_k^* - \underline{\varphi}_k^T(u)\ \hat{\underline{\theta}}\right)^2 \tag{1.3}$$

where

$$y_k^* = y_k + b_k \tag{1.4}$$

and

$$\varepsilon_k = y_k^* - \widehat{y}_k\left(\underline{\widehat{\theta}}, u\right) \tag{1.5}$$

with

– y_k^* : output measurement,
– y_k : true value of the output,
– b_k : stochastic perturbation,
– ε_k : residual.

Since the model is linear with respect to its parameters, the value $\underline{\theta}_{MC}$ of $\underline{\widehat{\theta}}$ which minimize J, is obtained analytically [LJU 87]:

$$\underline{\theta}_{MC} = \left(\sum_{k=1}^{K} \underline{\varphi}_k \underline{\varphi}_k^T\right)^{-1} \sum_{k=1}^{K} \underline{\varphi}_k y_k^* \tag{1.6}$$

Using (1.4), we obtain:

$$\underline{\theta}_{MC} = \underline{\theta} + \Delta\underline{\theta}_{MC} \tag{1.7}$$

where $\Delta\underline{\theta}_{MC}$ is the estimation error such that

$$\Delta\underline{\theta}_{MC} = \left(\sum_{k=1}^{K} \underline{\varphi}_k \underline{\varphi}_k^T\right)^{-1} \sum_{k=1}^{K} \underline{\varphi}_k b_k \tag{1.8}$$

Consider that the system is in open-loop, *i.e.* the sequence $\{u_k\}$ is independent of the noise sequence $\{b_k\}$. Moreover, let us suppose that the perturbation is zero-mean. Then:

$$E\left\{\underline{\varphi}_k b_k\right\} = 0 \tag{1.9}$$

and

$$E\left\{\Delta\underline{\theta}_{MC}\right\} = \lim_{K\to\infty} \Delta\underline{\theta}_{MC} = 0 \tag{1.10}$$

i.e. the estimate $\underline{\theta}_{MC}$ is asymptotically unbiased [LJU 87, WAL 97].

Let us suppose that the perturbation $\{b_k\}$ is independent and stationary, then

$$Var\{b_k\} = \sigma^2$$

The variance of the estimate $Var\{\underline{\theta}_{MC}\}$ is given by [LJU 87]:

$$Var\{\underline{\theta}_{MC}\} = \sigma^2 \left(\sum_{k=1}^{K} \underline{\varphi}_k \underline{\varphi}_k^T \right)^{-1} \tag{1.11}$$

Since the term σ^2 is unknown, it is necessary to estimate it:

$$\widehat{\sigma}^2 = \frac{J(\underline{\theta}_{MC})}{K - N} \tag{1.12}$$

where

– $J(\underline{\theta}_{MC})$ is the value of the criterion with $\widehat{\underline{\theta}} = \underline{\theta}_{MC}$,

– K is the number of measurements,

– N is the number of parameters.

REMARK. The expression (1.11) is attractive by its simplicity; unfortunately, this expression is rarely used because it is unrealistic. In fact, it is rare that the random perturbation verifies the preceding hypotheses. Generally, the perturbation can be correlated and non stationary. Moreover, a deterministic part corresponding to modeling error adds to the perturbation. This part is directly correlated to the input $\{u_k\}$. Then $E\left\{\underline{\varphi}_k m_k\right\} \neq 0$ and the estimator is asymptotically biased. Hence, (1.11) cannot be used to define realistic uncertainty domains of estimated parameters. Nevertheless, this relation gives valued information on the sensitivity of the quadratic criterion in comparison with the variation of the parameters (in the neighborhood of the optimum), which is a relative uncertainty (and not an absolute one).

1.2.3. *Principle of the Output-Error method in the general case*

Consider a system given by the general N^{th} order state-space model depending on parameter vector $\underline{\theta}$:

$$\begin{cases} \dot{\underline{x}} = \underline{g}(\underline{x}, \underline{\theta}, u) \\ y = f(\underline{x}, \underline{\theta}, u) \end{cases} \tag{1.13}$$

where $y(t)$ and $u(t)$ are considered as single output and input to simplify the presentation. One can notice that no hypothesis on the linearity is necessary: $\underline{g}$ and f are

based on the physical laws, which are generally non-linear. The only hypothesis is that the system is identifiable. [WAL 97].

Let $\widehat{\underline{\theta}}$ be an estimation of $\underline{\theta}$. Then, using the input $u(t)$ and a numerical integration technique, a simulation of the system output is obtained:

$$\begin{cases} \dot{\widehat{\underline{x}}} = \underline{g}\left(\widehat{\underline{x}}, \widehat{\underline{\theta}}, u\right) \\ \widehat{y} = f\left(\widehat{\underline{x}}, \widehat{\underline{\theta}}, u\right) \end{cases} \tag{1.14}$$

The optimum $\underline{\theta}_{opt}$ is the value which minimizes the quadratic criterion:

$$J = \sum_{k=1}^{K} \varepsilon_k^2 = \sum_{k=1}^{K} \left(y_k^* - \hat{f}_k\left(u, \hat{\underline{\theta}}\right)\right)^2 \tag{1.15}$$

where y_k^* is the measure of the noised output, b_k being the noise signal.

Since $\widehat{y}(t)$ is not linear in parameters $\widehat{\underline{\theta}}$, a non-linear optimization technique is used [RIC 71]. Many techniques can be used; our choice is based on techniques based on the gradient and more precisely on the Levenberg-Marquardt [MAR 63] algorithm:

$$\underline{\theta}_{j+1} = \underline{\theta}_j - \left\{\left[J''_{\theta\theta} + \lambda I\right]^{-1} \underline{J}'_{\theta}\right\}_{\hat{\underline{\theta}} = \underline{\theta}_j} \tag{1.16}$$

where

– $\underline{J}'_{\theta} = -2 \sum_{k=1}^{K} \varepsilon_k \, \underline{\sigma}_{k,\underline{\theta}_i}$: gradient,

– $J''_{\theta\theta} \approx 2 \sum_{k=1}^{K} \underline{\sigma}_{k,\underline{\theta}_i} \underline{\sigma}^T_{k,\underline{\theta}_i}$: Gauss-Newton's approximation of the Hessian,

– λ: control parameter,

– $\underline{\theta}_{k,\underline{\theta}_i} = \frac{\partial \hat{y}_k}{\partial \underline{\theta}_i}$: output sensitivity function.

This algorithm, using the control parameter λ, tends to the gradient algorithm when the estimation is far from the optimum (then $\lambda >> 1$). Near the optimum, it tends to a Newton's technique (when $\lambda \longrightarrow 0$) which allows acceleration of the convergence near the optimum. This procedure ensures robust convergence, even with a bad initialization.

Fundamentally, this technique is based on the calculation of gradient and Hessian, themselves dependent on the numerical integration of the sensitivity functions [TRI 88, KNU 94, WAL 97]. These sensitivity functions $\underline{\sigma}$ are equivalent to the regressor $\underline{\varphi}$ in the L.P. case. Thus, let us consider the simulation of $y(t)$ obtained with

the estimated parameters $\widehat{\underline{\theta}}$:

$$\hat{y}_k = f_k\left(\underline{\hat{x}}, \underline{\hat{\theta}}, u\right) \tag{1.17}$$

and let $d\widehat{\underline{\theta}}$ be a variation of $\widehat{\underline{\theta}}$. Then

$$\hat{y}_k\left(\underline{\hat{\theta}} + d\underline{\hat{\theta}}\right) = \hat{y}_k\left(\underline{\hat{\theta}}\right) + \underline{\sigma}^T_{k,\underline{\hat{\theta}}} d\underline{\hat{\theta}} + \cdots \tag{1.18}$$

or, with a development limited to the first order:

$$d\hat{y}_k \approx \underline{\sigma}^T_{k,\underline{\hat{\theta}}} d\underline{\hat{\theta}} \tag{1.19}$$

In the L.P. case, the prediction $\widehat{y}_k$ is given by:

$$\hat{y}_k = \underline{\varphi}^T_{k,\underline{\hat{\theta}}} \underline{\hat{\theta}} \tag{1.20}$$

or

$$d\hat{y}_k = \underline{\varphi}^T_{k,\underline{\hat{\theta}}} d\underline{\hat{\theta}} \tag{1.21}$$

Moreover, if $\underline{\theta}_{opt}$ is the parameter vector which minimizes J, and if we consider a variation $d\widehat{\underline{\theta}}$ around $\underline{\theta}_{opt}$, it is easy to show that:

$$J\left(\underline{\theta}_{opt} + d\underline{\hat{\theta}}\right) \approx J\left(\underline{\theta}_{opt}\right) + d\underline{\hat{\theta}}^T \left(\sum_{k=1}^{K} \underline{\sigma}_k \underline{\sigma}^T_k\right) d\underline{\hat{\theta}} \tag{1.22}$$

In the L.P. case, if $\underline{\theta}_{MC}$ is the parameter vector which minimizes J, one obtains:

$$J\left(\underline{\theta}_{MC} + d\underline{\hat{\theta}}\right) = J\left(\underline{\theta}_{MC}\right) + d\underline{\hat{\theta}}^T \left(\sum_{k=1}^{K} \underline{\varphi}_k \underline{\varphi}^T_k\right) d\underline{\hat{\theta}} \tag{1.23}$$

These two examples show the analogy between regressors and sensitivity functions and the interest of the sensitivity functions to analyze the Output-Error estimators, particularly their accuracy.

1.2.4. *Sensitivity functions*

In practice, it is necessary to differentiate two kinds of sensitivity functions [KNU 94, MOR 99]:

– $\sigma_{y,\theta_i} = \frac{\partial \widehat{y}(t)}{\partial \theta_i}$: output sensitivity function, used for the calculation of the gradient and the hessian,

– $\sigma_{x_n,\theta_i} = \frac{\partial x_n(t)}{\partial \theta_i}$: state sensitivity function.

Let us recall that

$$\begin{cases} \dim(\underline{x}) = N \\ \dim(\underline{\theta}) = I \end{cases}$$

Then, $\sigma_{y,\underline{\theta}}$ is a vector of dimension I and $\sigma_{x_n,\underline{\theta}}$ is a matrix of dimension $[N \times I]$ such that

$$\underline{\sigma}_{\underline{x},\underline{\theta}} = \left[\begin{array}{cccccc} \underline{\sigma}_{\underline{x},\theta_1} & \cdots & \underline{\sigma}_{\underline{x},\theta_i} & \cdots & \underline{\sigma}_{\underline{x},\theta_I} \end{array} \right] \tag{1.24}$$

For each parameter θ_i, σ_{x_n,θ_i} is obtained by partial derivation of the equation (1.13). Thus:

$$\frac{\partial \dot{\underline{x}}}{\partial \theta_i} = \dot{\underline{\sigma}}_{\underline{x},\theta_i} = \frac{\partial g(\underline{x},\underline{\theta},u)}{\partial \underline{x}} \frac{\partial \underline{x}}{\partial \theta_i} + \frac{\partial g(\underline{x},\underline{\theta},u)}{\partial \theta_i} \tag{1.25}$$

Then, σ_{x_n,θ_i} is the solution of a non-linear differential system:

$$\dot{\underline{\sigma}}_{\underline{x},\theta_i} = \frac{\partial g(\underline{x},\underline{\theta},u)}{\partial \underline{x}} \underline{\sigma}_{\underline{x},\theta_i} + \frac{\partial g(\underline{x},\underline{\theta},u)}{\partial \theta_i} \tag{1.26}$$

Finally, $\partial y / \partial \theta_i$ is obtained by partial derivation of equation (1.13):

$$\frac{\partial y}{\partial \theta_i} = \left(\frac{\partial f(\underline{x},\underline{\theta},u)}{\partial \underline{x}} \right)^T \underline{\sigma}_{\underline{x},\theta_i} + \frac{\partial f(\underline{x},\underline{\theta},u)}{\partial \theta_i} \tag{1.27}$$

1.2.5. *Convergence of the estimator*

Because the simulation $\widehat{y} = f(\widehat{\underline{x}},\widehat{\underline{\theta}},u)$ is non-linear in $\widehat{\underline{\theta}}$, the quadratic criterion $J(\widehat{\underline{\theta}})$ is not parabolic as in the L.P. case and the uniqueness of the optimum $\underline{\theta}_{opt}$ is not guaranteed [WAL 97]. Secondary optima can exist and the optimization algorithm can converge to one of these optima. A solution to converge to the global optimum consists to use a global approach, like genetic algorithms [FON 95]. A "map" of the criterion and of its optima is obtained. The global optimum $\underline{\theta}_{opt}$ is then obtained using the Marquardt's algorithm.

In the presence of noise affecting the output, the estimator converges to:

$$\underline{\theta}_{opt} = \underline{\theta} + \Delta\underline{\theta} \tag{1.28}$$

where $\underline{\theta}$ is the true parameter vector and $\Delta\underline{\theta}$ the estimation error. Using the sensitivity functions, $\Delta\underline{\theta}$ can be approached by:

$$\Delta\underline{\theta} \approx \left\{ \left(\sum_{k=1}^{K} \underline{\sigma}_k \underline{\sigma}_k^T \right)^{-1} \left(\sum_{k=1}^{K} \underline{\sigma}_k b_k \right) \right\}_{\underline{\theta}_{opt}} \tag{1.29}$$

which is equivalent to:

$$\Delta\underline{\theta} = \left\{ \left(\sum_{k=1}^{K} \underline{\varphi}_k \underline{\varphi}_k^T \right)^{-1} \left(\sum_{k=1}^{K} \underline{\varphi}_k b_k \right) \right\}_{\underline{\theta}_{opt}} \tag{1.30}$$

obtained in the L.P. case.

Then, if $E\{b_k\} = 0$ and if the system is in open-loop, u_k is independent of b_k and $E\{\Delta\underline{\theta}\} = 0$ (since $\underline{\sigma}_k$ depends only on u_k). The Output-Error estimator is asymptotically unbiased, whatever the nature of the zero mean output noise [WAL 97].

REMARK. When the system is in closed-loop (which is the case with AC machines with vector control), the perturbation $\{b_k\}$ is necessarily correlated with the control input through the corrector (or the control algorithm). Then, $\underline{\sigma}_k$ which depends of the input $\{u_k\}$ is correlated with the noise $\{b_k\}$, *i.e.* the estimation $\underline{\theta}_{opt}$ is asymptotically biased. Fortunately, this bias is really very significant only when the signal to noise ratio is weak. Thus, as a first approximation, it can be neglected. Moreover, Output-Error algorithms functioning in closed-loop have been proposed; they are able to reject this bias using a more complicated identification procedure [GRO 00, LAN 97].

1.2.6. *Variance of the estimator*

Using the analogy between regressor and sensitivity functions, it is possible to define an approximated expression of the variance of $\underline{\theta}_{opt}$. Thus, by replacing $\underline{\varphi}_k$ by $\underline{\sigma}_k$ in expression (1.11), one obtains:

$$Var\left\{\underline{\theta}_{opt}\right\} \approx \hat{\sigma}^2 \left(\sum_{k=1}^{K} \underline{\sigma}_k \underline{\sigma}_k^T \right)^{-1} \tag{1.31}$$

Note that this expression is approximated for the following reasons:

– since the output model is non linear in parameters, $\underline{\sigma}_k$ is a local approximation of $\underline{\varphi}_k$, *i.e.* this expression is an approximation of the quadratic criterion by a paraboloid;

– like in the L.P. case, the perturbation is not characterized by the only term $\widehat{\sigma}^2$, *i.e.* this expression gives only an information on the relative accuracy.

1.2.7. *Implementation*

Output-Error algorithms are more complex than Equation-Error algorithms because of the non-linear optimization. However, this is not the only difficulty to implement these algorithms. It is also necessary to study the numerical simulation, the problem of initial conditions and the normalization of the sensitivity functions.

1.2.7.1. *Simulation of the differential system*

In the Least-Squares case, the implementation of the predictor is trivial in the discrete-time case; in the continuous-time case, it is more difficult [TRI 01].

In the Output-Error case, we are confronted with a real problem of numerical simulation of differential systems, for the simulation of the model output and the sensitivity functions. Simulation error or approximate calculations will give a systematic error or deterministic bias.

When the system is linear, the differential equations can be integrated using the exponential matrix technique [ISE 92] which allows conciliation of accuracy and rapidity of the computations.

When the system is non-linear, which is almost always the case with physical systems, it is necessary to give importance to accuracy and numerical stability. Accuracy is guaranteed by an algorithm like Runge-Kutta of order 4. On the other hand, the discretization of the differential equations can make their integration unstable. Then, implicit techniques can be used, for instance the Adams' algorithms [NOU 89].

Nevertheless, some errors can subsist, even if all the preceding precautions are taken. In fact, it is very important to well consider the type of the input in the numerical integration algorithm: the simplest case is when the input is derived from a numerical command. The applied input is then continuous and it is necessary to specify how the input varies between two sample instants: a linear change is sufficient in the major cases, otherwise, a parabolic (or of a higher order) extrapolation should be used [ISE 92].

1.2.7.2. *Initial conditions in OE algorithms*

Initial conditions of the state are supplementary unknown parameters which need identification [WAL 97]. When the number of data is relatively small, particularly

in regard of the transient of the system output, that is the necessary solution. Then, an extended parameter vector is considered which contains the system parameters and the initial conditions. The identification algorithm remains unchanged, but the computation time and the convergence difficulties increase. When the acquisition time is higher than the transient t_r [RIC 91], the k_r first data ($k_r = t_r/T_e$) are not used in the criterion in order to avoid the estimation of initial conditions. Then, the criterion becomes:

$$J = \sum_{k=k_r}^{K} \varepsilon_k^2 \tag{1.32}$$

and the differential system is simulated from $t = 0$.

Let us recall that in the two situations, if the initial conditions are not taken into account, an important bias appears.

1.2.7.3. *The normalization*

In an academic situation, the user chooses numerical values which are close. In a real situation, numerical values can be very different: then, some difficulties for the convergence of the identification algorithm appear. A solution consists to normalize the parameters, which in practice consists in the normalization of the sensitivity functions [MOR 99, RIC 71].

Consider the parameter θ_n, with an initial estimation θ_{n_0}, such that $\theta_n = \theta_{n_0} + \Delta\theta_n$. The estimation of θ_n is equivalent to that of $\Delta\theta_n$.

Let us define $\theta_n = (1 + \mu_n)\theta_{n_0}$ where $\Delta\theta_n = \mu_n\theta_{n_0}$. The sensitivity function linked to θ_n is then :

$$\frac{\partial\hat{y}}{\partial\theta_n} = \frac{1}{\theta_{n_0}}\frac{\partial\hat{y}}{\partial\mu_n} \tag{1.33}$$

where the sensitivity functions $\partial\widehat{y}/\partial\mu_n$ are now normalized and closed.

In practice, $\underline{\mu}$ and thus $\underline{\theta}$ are estimated using the non linear programming algorithm and we obtain:

$$\theta_n = (1 + \mu_n)\,\theta_{n_0} \tag{1.34}$$

1.3. Parameter estimation with *a priori* information

1.3.1. *Introduction*

Despite all the numerical precautions stated in the preceding paragraph, Output-Error algorithms can, in certain situations, provide incoherent estimates such as negative resistances or inductances (refer to [ALM 95, JEM 97] in the electrical engineering case). It is necessary to seek the cause of these anomalies in the optimization

mechanism: indeed, the latter determines the set of parameters which allows best fit of the data by the selected model, without any physical constraint. Moreover, when that occurs, only some parameters are affected by this anomaly.

It is fundamentally a problem of parameter sensitization: although theoretically identifiable, the concerned parameters are almost unidentifiable and balancing phenomena can be observed. The traditional reflex in front of such problem is to propose to improve the excitation [LJU 87, KAB 97]: however, in many situations, this optimal excitation can turn out to be unrealistic in practice (even dangerous for the process) or can transgress the validity conditions of the model!

In addition, these problems of excitation must be connected to the selected model and its modeling error. Within the framework of control using black-box models, the engineer increases the complexity of these models until the residuals become independent and uncorrelated to the input [LJU 87]. On the other hand, the "physicist" uses models of voluntarily reduced complexity, adapted to the description of a specific phenomenon: that does not prevent him to estimate the corresponding parameters without systematic error, but with specific and dedicated approaches. Then, the question is to know if these same models, used with the identification techniques developed in automatic control, can provide estimates which are coherent with physical laws.

A solution suggested in this chapter consists in introducing explicitly the physical knowledge in order to replace the lack of excitation or to improve it.

In addition, even with a good excitation, it is interesting to improve the convergence of the estimator (and its accuracy) by introducing an *a priori* knowledge (given that this knowledge is not erroneous!).

1.3.2. *Bayesian approach*

A possible misunderstanding should be immediately cleared up: effectively, it is recommended to initialize the optimization algorithm with parameters close to the optimum in order to accelerate the convergence and to reduce the computation time. For this, an initial knowledge is used but the optimization algorithm quickly forgets this initialization! In the best situation, the convergence to a secondary optimum can be avoided.

It is therefore necessary to introduce explicitly this prior knowledge: thus, a modified or compound criterion is defined. The major justification of this new criterion can be found in the Bayesian approach [EYK 74, GOO 77, PET 81, TUL 93]. This approach consists in considering the estimation problem in a probabilistic context.

Consider a set of experimental data u_k, y_k^* (or $\underline{U}, \underline{Y}^*$); we propose to estimate $\underline{\theta}$ by maximizing the probability density of $\widehat{\underline{\theta}}$ conditionally to the data $\underline{Y}^*$, *i.e.* $P\widehat{\underline{\theta}}/\underline{Y}^*$ (or

a posteriori density). Otherwise, we have *a priori* information on $\widehat{\underline{\theta}}$ characterized by $P\widehat{\underline{\theta}}$. Then, the Bayes's relation is:

$$P\underline{\hat{\theta}}\Big/\underline{Y}^* = \frac{P\underline{\hat{\theta}}P\underline{Y}^*\Big/\underline{\hat{\theta}}}{P\underline{Y}^*} \tag{1.35}$$

Because $P\underline{Y}^*$ does not explicitly depend on $\widehat{\underline{\theta}}$, the maximization of $P\widehat{\underline{\theta}}/\underline{Y}^*$ is equivalent to the maximization of $P\widehat{\underline{\theta}}P\underline{Y}^*/\widehat{\underline{\theta}}$: this is the technique known as *a posteriori* maximum.

In order to tackle this problem, some hypothesis are necessary: usually, $P\widehat{\underline{\theta}}$ and $P\underline{Y}^*/\widehat{\underline{\theta}}$ are considered as Gaussian densities. Then, we can write:

$$\begin{aligned} P\underline{\hat{\theta}}\Big/\underline{Y}^* = \quad & A\exp\left[-\tfrac{1}{2}\left(\underline{\hat{\theta}}-\underline{\theta}_0\right)^T M_0^{-1}\left(\underline{\hat{\theta}}-\underline{\theta}_0\right)\right. \\ & \left.-\tfrac{1}{2}\left(\underline{Y}^*-\underline{\hat{Y}}\left(\underline{\hat{\theta}},\underline{U}\right)\right)^T R_b^{-1}\left(\underline{Y}^*-\underline{\hat{Y}}\left(\underline{\hat{\theta}},\underline{U}\right)\right)\right] \end{aligned} \tag{1.36}$$

where:

– A is a constant,

– $\underline{\theta}_0$: initial knowledge of $\underline{\theta}$,

– M_0 : covariance matrix of $\underline{\theta}_0$,

– R_b : covariance matrix of the stochastic perturbation.

Finally, by considering the natural logarithm of this expression, it is easy to show that the maximization of $P\widehat{\underline{\theta}}/\underline{Y}^*$ is equivalent to the minimization of the compound criterion:

$$\begin{aligned} J_C = \quad & \underbrace{\left(\underline{\hat{\theta}}-\underline{\theta}_0\right)^T M_0^{-1}\left(\underline{\hat{\theta}}-\underline{\theta}_0\right)}_{J_0} \\ & +\underbrace{\left(\underline{Y}^*-\underline{\hat{Y}}\left(\underline{\hat{\theta}},\underline{U}\right)\right)^T R_b^{-1}\left(\underline{Y}^*-\underline{\hat{Y}}\left(\underline{\hat{\theta}},\underline{U}\right)\right)}_{J^*} \\ J_C = \quad & J_0+J^* \end{aligned} \tag{1.37}$$

In practice, the covariance matrix of the perturbation is unknown and R_b is replaced by $\sigma_b^2 I$ (where σ_b^2 is the variance of the noise and I the identity matrix). Then, except σ_b^2 (representing the variance of output noise), J^* represents the usual quadratic criterion, containing experimental information. On the other hand, J_0 is a second quadratic criterion which introduces an "elastic" constraint in the minimization of the global criterion J_C: in fact, it prevents $\widehat{\underline{\theta}}$ to move away from $\underline{\theta}_0$, with a "return force" dependent on $\widehat{\underline{\theta}}-\underline{\theta}_0$.

1.3.3. *Minimization of the compound criterion*

Let

$$J_C = \left(\underline{\hat{\theta}} - \underline{\theta}_0\right)^T M_0^{-1} \left(\underline{\hat{\theta}} - \underline{\theta}_0\right) + \frac{1}{\sigma_b^2} \sum_{k=1}^{K} \left(y_k^* - \hat{y}_k \left(\underline{\hat{\theta}}, u\right)\right)^2 \tag{1.38}$$

This new criterion is minimized using the Marquardt's algorithm with:

$$\underline{J}'_{C_{\underline{\theta}}} = 2 \left[M_0^{-1} \left(\underline{\hat{\theta}} - \underline{\theta}_0\right) - \frac{1}{\hat{\sigma}_b^2} \sum_{k=1}^{K} \varepsilon_k \underline{\sigma}_k \right] \tag{1.39}$$

and

$$J''_{C_{\theta\theta}} \approx 2 \left[M_0^{-1} + \frac{1}{\hat{\sigma}_b^2} \sum_{k=1}^{K} \underline{\sigma}_k \underline{\sigma}_k^T \right] \tag{1.40}$$

Let $\underline{\theta}_C$ be the value of $\widehat{\underline{\theta}}$ minimizing J_C and obtained using the optimization algorithm. In order to demonstrate the interest of *a priori* information, we propose to show the links between $\underline{\theta}_C$, $\underline{\theta}_{opt}$ and $\underline{\theta}_0$.

Thus, let us consider that $\underline{\theta}_{opt}$ and $\underline{\theta}_0$ are near to $\underline{\theta}_C$, *i.e.* the sensitivity functions $\widehat{\underline{\theta}}$ are approximately equal. Then, one can write:

$$J_C \left(\underline{\hat{\theta}}\right) \approx J_{C\,\min} + \left(\underline{\hat{\theta}} - \underline{\theta}_C\right)^T \left(J''_{\theta\theta}\right)^{-1} \left(\underline{\hat{\theta}} - \underline{\theta}_C\right) \tag{1.41}$$

where

$$\underline{J}'_{\theta} = 2 \left(M_0^{-1} \left(\underline{\hat{\theta}} - \underline{\theta}_0\right) - \frac{S^T \underline{\varepsilon}}{\sigma_b^2} \right) \tag{1.42}$$

and

$$J''_{\theta\theta} \approx 2 \left(M_0^{-1} + \frac{S^T S}{\sigma_b^2} \right) \tag{1.43}$$

with

$$S = \begin{bmatrix} \underline{\sigma}_1^T \\ \vdots \\ \underline{\sigma}_K^T \end{bmatrix}$$

When the Newton's algorithm is used to minimize J_C from the initial value $\underline{\theta}_0$, this optimum $\underline{\theta}_C$ is obtained in one iteration because J_C is a quadratic form of $\widehat{\underline{\theta}}$. Then:

$$\underline{\theta}_C = \underline{\theta}_0 - \left\{ \left[J''_{\theta\theta} \right]^{-1} \underline{J}'_{\theta} \right\}_{\underline{\theta}_0} \tag{1.44}$$

with

$$\left\{ \underline{J}'_{\theta} \right\}_{\underline{\theta}_0} = \frac{-2S^T}{\sigma_b^2} \left(\underline{Y}^* - \underline{\hat{Y}} \left(\underline{\theta}_0 \right) \right) \tag{1.45}$$

which gives:

$$\underline{\theta}_C = \underline{\theta}_0 + \left[M_0^{-1} + \frac{S^T S}{\sigma_b^2} \right]^{-1} \frac{S^T}{\sigma_b^2} \left(\underline{Y}^* - \underline{\hat{Y}} \left(\underline{\theta}_0 \right) \right) \tag{1.46}$$

Let $P_{opt}^{-1} = \frac{S^T S}{\sigma_b^2}$ where P_{opt} is the covariance matrix linked to the conventional criterion minimized by $\underline{\theta}_{opt}$. Then, let us define:

$$P^{-1} = M_0^{-1} + P_{opt}^{-1} \tag{1.47}$$

i.e.

$$\underline{\theta}_C = \underline{\theta}_0 + \frac{PS^T}{\sigma_b^2} \left(\underline{Y}^* - \underline{\hat{Y}} \left(\underline{\theta}_0 \right) \right) \tag{1.48}$$

which can be written as:

$$\underline{\theta}_C = \underline{\theta}_0 + K \left(\underline{Y}^* - \underline{\hat{Y}} \left(\underline{\theta}_0 \right) \right) \tag{1.49}$$

where K is the gain of a Kalman filter applied to the particular system [RAD 84]:

$$\begin{cases} \underline{\theta}_{i+1} = \underline{\theta}_i = \underline{\theta} \\ \underline{Y}_i^* = \underline{Y}_i \left(\underline{\theta}, u \right) + \underline{b}_i \end{cases} \tag{1.50}$$

This interpretation confirms the preceding probabilistic and stochastic approach: from $\underline{\theta}_0$, the optimal value to obtain $\underline{\theta}_C$ is determined, taking into account the gain K and the information given by the data set $\underline{Y}^*$ (corresponding to $\underline{\theta}_{opt}$). Moreover, taking into account [1.47], the estimate $\underline{\theta}_C$ is necessarily more precise than $\underline{\theta}_0$ or $\underline{\theta}_{opt}$.

1.3.4. *Deterministic interpretation*

The Bayesian approach is comforted by the interpretation of the Kalman filter which can be interpreted like a regularization technique [TIK 77, JOH 97, SAY 98]. Unfortunately, this interpretation failed in the case of physical parameter estimation using a model of reduced complexity.

It is to be noted that the output perturbation $\{b_k\}$ is not really a random perturbation but a deterministic one: the major part of the output perturbation is due to the modeling error. The perturbation is determinist because it is generated by the input u.

Thus, a new interpretation is proposed, essentially based on the geometry of the quadratic criterion (see [MOR 99] for more information).

Let $\widehat{\sigma}_b^2$ be the pseudo-variance of the output perturbation which is obtained using:

$$\widehat{\sigma}_b^2 = \frac{J(\underline{\theta}_{opt})}{K} \tag{1.51}$$

where J is the conventional criterion and K the number of samples.

We can notice that $J\left(\underline{\theta}_{opt}\right) = \sum_{k=1}^{K} \left(y_k^* - f_k\left(u_k, \underline{\theta}_{opt}\right)\right)^2$ takes into account the noise b_k and the modeling error, *i.e.* $\widehat{\sigma}_b^2$ is a pseudo-variance.

Using $\widehat{\sigma}_b^2$, a pseudo-covariance matrix can be defined by:

$$P_{opt} = \hat{\sigma}_b^2 \left(S^T S\right)^{-1} \tag{1.52}$$

In the neighborhood of $\underline{\theta}_{opt}$, it is possible to write:

$$\frac{J\left(\hat{\underline{\theta}}\right)}{\hat{\sigma}_b^2} \approx \left(\hat{\underline{\theta}} - \underline{\theta}_{opt}\right)^T \frac{S^T S}{\hat{\sigma}_b^2} \left(\hat{\underline{\theta}} - \underline{\theta}_{opt}\right) + \frac{J_{opt}}{\hat{\sigma}_b^2} \tag{1.53}$$

Finally:

$$J_C \approx \left(\hat{\underline{\theta}} - \underline{\theta}_0\right)^T M_0^{-1} \left(\hat{\underline{\theta}} - \underline{\theta}_0\right) + \left(\hat{\underline{\theta}} - \underline{\theta}_{opt}\right)^T P_{opt}^{-1} \left(\hat{\underline{\theta}} - \underline{\theta}_{opt}\right) + K \tag{1.54}$$

This expression shows that J_C is the result of two paraboloid, centered on $\underline{\theta}_0$ and $\underline{\theta}_{opt}$, characterized by the spreads M_0 and P_{opt}. J_C is a unique paraboloid, centered on $\underline{\theta}_C$, such that:

$$\underline{\theta}_C \approx \underline{\theta}_0 + \left[M_0^{-1} + P_{opt}^{-1}\right]^{-1} \left(M_0^{-1}\underline{\theta}_0 + P_{opt}^{-1}\underline{\theta}_{opt}\right) \tag{1.55}$$

($\underline{\theta}_C$ is the barycentre of $\underline{\theta}_0$ and $\underline{\theta}_{opt}$) characterized by the spread M_C, such that:

$$M_C^{-1} = M_0^{-1} + P_{opt}^{-1} \tag{1.56}$$

Let

$$J_C \approx \left(\hat{\underline{\theta}} - \underline{\theta}_C\right)^T M_C^{-1} \left(\hat{\underline{\theta}} - \underline{\theta}_C\right) + J_C\left(\underline{\theta}_C\right) \tag{1.57}$$

Let us consider a monovariable example in order to illustrate this result (figure 1.1):

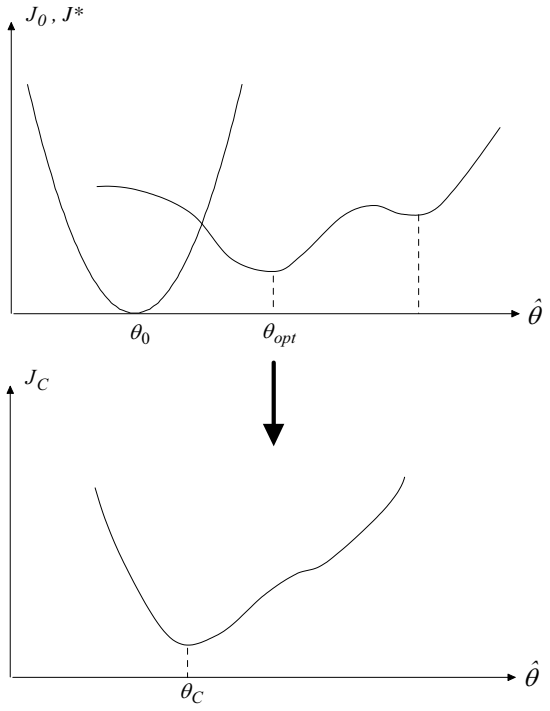

Figure 1.1. *Determinist interpretation of the compound criterion - monovariable case*

When $\underline{\theta}_0$ is close to $\underline{\theta}_{opt}$, the criterion J_C is more convex than the conventional criterion. Secondary optima tend to disappear. Consequently, the introduction of

$\{\underline{\theta}_0, M_0\}$ accelerates the convergence of the optimization algorithm. The estimation $\underline{\theta}_C$ is generally more precise than $\underline{\theta}_0$ or $\underline{\theta}_{opt}$ because the paraboloid J_C is less "open" than the preceding criteria.

Moreover, if the estimation $\underline{\theta}_{opt}$ is ill-sensitized, *i.e.* the optimum is straight, then the global optimum $\underline{\theta}_C$ will be attracted by $\underline{\theta}_0$ which is equivalent to an "elastic" return force (especially if the paraboloid J_0 is "closed").

REMARK. The conventional probabilistic interpretation has certainly been an obstacle to the use of the Bayesian approach. On the contrary, the determinist interpretation shows the advantages of this technique on the convergence of the optimization algorithm and on the uniqueness of the optimum.

Moreover, it is important to notice that the covariance matrices of $\underline{\theta}_0$ and of the noise can be significantly simplified, with no influence on their properties. Thus, matrix M_0 can be favorably reduced to a diagonal form, with an overevaluation of each variance if necessary as it will be shown in the applications paragraph. Moreover, the covariance matrix of the noise can be reduced to a single coefficient constituted by the variance of the perturbation, which is particularly justified when a modeling error occurs.

1.3.5. *Implementation*

The implementation of this methodology implies handling of two types of information:

– the *a priori* information $\{\underline{\theta}_0, M_0\}$ obtained from a preceding global estimation or from specific and partial estimations. In this more realistic case, the matrix M_0 is at best diagonal (some examples are proposed in the next paragraph);

– the variance $\widehat{\sigma}_b^2$ of the perturbation: this is an essential parameter in the weighting between J_0 and J. The deterministic interpretation has shown that $\widehat{\sigma}_b^2$ is linked to the spread of the paraboloid $J/\widehat{\sigma}_b^2$: a lower value gives more weight to experimental data, while a high value gives more importance to the initial information $\underline{\theta}_0$.

In practice, it is necessary to define a practical procedure for its determination because it principally depends on the modeling error, unknown before optimization. Concretely, a prior value is used, which is approximately the variance of the stochastic perturbation; when $\underline{\theta}_C$ is estimated, a new value $\widehat{\sigma}_b^2$ is obtained using $J(\underline{\theta}_C)$ (where J is the conventional criterion). If $\widehat{\sigma}_b^2$ is very far from its initial value, a new estimation of $\underline{\theta}_C$ is performed. This last point is considered in the next paragraph.

The value of $\widehat{\sigma}_b^2$ obtained using $J\left(\underline{\theta}_C\right)$ can also be used as a guideline of the coherence between initial information and prior knowledge. Thus, when $\widehat{\sigma}_b^2$, $\underline{\theta}_0$ and

M_0 are reliable information, $\widehat{\sigma}_b^2$ allows us to test if the experimental information is compatible with the knowledge on the system: a value of $\widehat{\sigma}_b^2$ significantly greater than σ_b^2 can correspond to a non-stationarity of the system which involves a modification of its model. Then, the modeling error increases and $\widehat{\sigma}_b^2$ too.

REMARK. M_0 and $\widehat{\sigma}_b^2$ play an essential part in the weighting between *a priori* information and experimental data. The variance of the *a priori* information M_0 should result from a decisive choice: a low variance increases confidence in $\underline{\theta}_0$. Thus, in practice, this value can be increased in order to allow the variation of $\underline{\theta}_C$, if this variation is expectable (see the paragraph dedicated to fault diagnosis).

The estimated value $\widehat{\sigma}_b^2$ can need to be adjusted thanks to a new identification, as it was previously described : in practice, the example treated in the next paragraph shows that this convergence is very fast.

Finally, let us recall that the use of the Bayesian approach must be justified by a true *a priori* information $\{\underline{\theta}_0, M_0\}$: it is necessary to be persuaded that an erroneous initial information will irremediably bias the estimator. However, when this use is justified, this technique improves significantly the convergence of the optimization algorithm (and thus computing time) and guarantees the existence of only one optimum. In addition, it allows reconsideration of some techniques based on parameter estimation, like fault detection.

1.4. Parameter estimation of the induced machine

1.4.1. *Introduction*

We propose to illustrate the application of Output-Error techniques (with and without *a priori* knowledge) to the induced machine in the dynamic case.

At first, it is necessary to specify the model of this machine, firstly in the three-phase context and secondly using the Park's transformation, which is well adapted for parameter estimation of that kind of machine. Then, we present the model identification starting from experimental data with the two types of algorithms previously defined.

1.4.2. *Modeling in the three-phases referential*

The general model of the induced machine is obtained by considering a uniform air-gap and a sinusoidal distribution of the magnetic field. The machine is supposed to work in an unsaturated magnetic state and the iron losses are neglected. In those

conditions, the dynamic model of the machine with the leaks added up at the stator can be described by:

$$\begin{cases} \underline{u}_s &= [R_s]\,\underline{i}_s + \frac{d}{dt}\underline{\phi}_s \\ \underline{0} &= [R_r]\,\underline{i}_r + \frac{d}{dt}\underline{\phi}_r \\ \underline{\phi}_s &= [L_s]\,\underline{i}_s + [M_{sr}]\,\underline{i}_r \\ \underline{\phi}_r &= [M_{sr}]^T\,\underline{i}_s + [L_r]\,\underline{i}_r \end{cases} \tag{1.58}$$

with:

$$[R_s] = R_s \cdot I \quad \text{et} \quad [R_r] = R_r \cdot I$$

$$[L_s] = \begin{pmatrix} L_p + L_{fsa} & -\frac{L_p}{2} & -\frac{L_p}{2} \\ -\frac{L_p}{2} & L_p + L_f & -\frac{L_p}{2} \\ -\frac{L_p}{2} & -\frac{L_p}{2} & L_p + L_f \end{pmatrix}$$

$$[L_r] = \begin{pmatrix} L_p & -\frac{L_p}{2} & -\frac{L_p}{2} \\ -\frac{L_p}{2} & L_p & -\frac{L_p}{2} \\ -\frac{L_p}{2} & -\frac{L_p}{2} & L_p \end{pmatrix}$$

$$[M_{sr}] = \begin{pmatrix} L_p\cos(\theta) & L_p\cos(\theta + \frac{2\pi}{3}) & L_p\cos(\theta - \frac{2\pi}{3}) \\ L_p\cos(\theta - \frac{2\pi}{3}) & L_p\cos(\theta) & L_p\cos(\theta + \frac{2\pi}{3}) \\ L_p\cos(\theta + \frac{2\pi}{3}) & L_p\cos(\theta - \frac{2\pi}{3}) & L_p\cos(\theta) \end{pmatrix}$$

where:

– $\underline{u}_s$, $\underline{i}_s$ and $\underline{i}_r$ respectively represent the voltage vector, the stator currents vector and the rotor currents vector;

– $\underline{\phi}_s$ and $\underline{\phi}_r$: vectors of stator and rotor fluxes;

– R_s (resp. R_r) : resistance of a stator phase (resp. rotor);

– L_p and L_f: principal inductance and leakage inductance added up at the stator;

– θ : electrical angle of the position of the rotor.

The equations of voltages obtained above are relatively simple, even if the number of state-variables is high. Nevertheless, the matrix of mutual inductances $[M_{sr}]$ depends on the electrical angle θ. Then, the writing of the state-space representation remains complex. These equations are simplified using the Park's transformation [CHA 87, CAR 95, GRE 97] which is described below.

1.4.3. *Park's Transformation*

The Park's Transformation, largely used for the modeling of AC machines, corresponds to a projection of the three-phase variables on a turning diphasic frame, the goal being to eliminate the position in the matrices of mutual inductances. For that, it is sufficient to carry out a transformation of the three-phase system abc to the diphasic system $\alpha\beta$ using the transformation T_{23} (Concordia's transformation) followed by a rotation of reference frame $P(-\theta)$ in order to link the reference frame to the rotation Park's axes dq (Fig. 1.2).

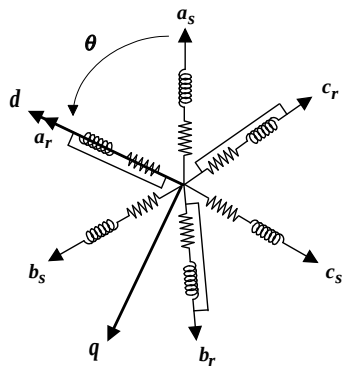

Figure 1.2. *Park's transformation linked to the rotor*

In a matrix way, the essential variables of the machine become [CAR 95]:
– in the stator : $\underline{x}_{dq_s} = P(-\theta)\, T_{23}\, \underline{x}_s$

– in the rotor : $\underline{x}_{dq_r} = T_{23}\, \underline{x}_r$

with

$$T_{23} = \sqrt{\frac{2}{3}} \begin{bmatrix} \cos(0) & \cos(\frac{2\pi}{3}) & \cos(\frac{4\pi}{3}) \\ \sin(0) & \sin(\frac{2\pi}{3}) & \sin(\frac{4\pi}{3}) \end{bmatrix}$$

$$P(\theta) = \begin{bmatrix} \cos(\theta) & \cos(\theta + \frac{\pi}{2}) \\ \sin(\theta) & \sin(\theta + \frac{\pi}{2}) \end{bmatrix} \text{: rotation matrix of angle } \theta$$

Equations of the induced machine [1.58] in the Park's frame linked to the rotor, with leaks added up at the stator are:

$$\begin{cases} \underline{U}_{dq_s} &= R_s\,\underline{i}_{dq_s} + \frac{d}{dt}\underline{\phi}_{dq_s} + \omega\,P(\frac{\pi}{2})\,\underline{\phi}_{dq_r} \\ \underline{0} &= R_r\,\underline{i}_{dq_r} + \frac{d}{dt}\underline{\phi}_{dq_r} \\ \underline{\phi}_{dq_s} &= (L_m + L_f)\,\underline{i}_{dq_s} + L_m\,\underline{i}_{dq_r} \\ \underline{\phi}_{dq_r} &= L_m\,(\underline{i}_{dq_s} + \underline{i}_{dq_r}) \end{cases} \tag{1.59}$$

where

– $\omega = \frac{d\theta}{dt}$ represents the electrical pulsation (where $\theta = p\,\theta_{mechanical}$ and p : number of pairs of poles per phase),

– $L_m = \frac{3}{2}L_p$: magnetizing inductance.

The obtained model of the induced machine is essentially characterized by four physical parameters R_s, R_r, L_m and L_f. These parameters are the parameters that need estimation.

1.4.4. *Continuous-time state-space model*

For the majority of the industrial applications of the induced machine, the inertia of the rotating parts is significant. Consequently, the rotor speed is generally slowly variable as compared to other electrical parameters of the machine [MOR 99]. Thus, a 4^{th} non-linear state-space representation of the induced machine is obtained (because of dependency on the speed) by associating the state-vector which contains the stator currents and rotor fluxes as well as the input and the output of the system corresponding respectively to the voltages and stator currents of axis d and q [CAR 95, MOR 99]:

$$\begin{cases} \dot{\underline{x}}(t) &= A(\omega)\,\underline{x}(t) + B\,\underline{u}(t) \\ \underline{y}(t) &= C\,\underline{x}(t) \end{cases} \tag{1.60}$$

with

$$\underline{x} = \begin{bmatrix} i_{d_s} & i_{q_s} & \phi_{d_r} & \phi_{q_r} \end{bmatrix}^T \text{ : state-vector}$$

$$\underline{u} = \begin{bmatrix} U_{d_s} \\ U_{q_s} \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} i_{d_s} \\ i_{q_s} \end{bmatrix} \text{ : respectively machine input and output}$$

$$A(\omega) = \begin{bmatrix} -\frac{R_s+R_r}{L_f} & \omega & \frac{R_r}{L_m \cdot L_f} & \frac{\omega}{L_f} \\ -\omega & -\frac{R_s+R_r}{L_f} & -\frac{\omega}{L_f} & \frac{R_r}{L_m \cdot L_f} \\ R_r & 0 & -\frac{R_r}{L_m} & 0 \\ 0 & R_r & 0 & -\frac{R_r}{L_m} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{L_f} & 0 & 0 & 0 \\ 0 & \frac{1}{L_f} & 0 & 0 \end{bmatrix}^T, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

1.4.5. *Output-Error identification*

The considered system is multivariable, with two inputs (U_{d_s} and U_{q_s}) and two outputs (i_{d_s} and i_{q_s}).

Thus, a criterion J composed of two quadratic terms is considered:

$$J = \sum_{k=1}^{K} \left(i^*_{d_{s_k}} - \hat{i}_{d_{s_k}} \right)^2 + \sum_{k=1}^{K} \left(i^*_{q_{s_k}} - \hat{i}_{q_{s_k}} \right)^2 \tag{1.61}$$

where $i^*_{d_{s_k}}$ and $i^*_{q_{s_k}}$ are sampled measurements with sampling period $T_e = 0.7\,ms$ ($t = kT_e$, k varying from 1 to $K = 4500$). $\hat{i}_{d_{s_k}}$ and $\hat{i}_{q_{s_k}}$ represent the simulation of the model [1.60] based on the estimation $\underline{\widehat{\theta}}$ where $\underline{\theta}^T = \left[\begin{array}{cccc} R_s & R_r & L_m & L_f \end{array} \right]$.

At each iteration, it is also necessary to simulate the sensitivity functions $\frac{\partial i_{d_s}}{\partial \underline{\widehat{\theta}}}$ and $\frac{\partial i_{q_s}}{\partial \underline{\widehat{\theta}}}$ according to the preceding paragraph.

Experimental data are obtained with an induced machine of 1.1 kW supplied by a generator with a vector control. The machine is regulated to its nominal speed, and coupled with a continuous generator which acts as a load.

The machine input is a pseudo-random binary sequence (PRBS) of ± 90 tr/mn added to the speed reference of 750 tr/mn. Rotor currents and voltages are measured, as well as the mechanical position of the rotor (which allows computation of the pulsation ω). The input vector $\{U_{d_s}, U_{q_s}\}$ and the output vector $\{i^*_{d_s}, i^*_{q_s}\}$ are obtained by using Park's transformation.

Minimizing the quadratic criterion (1.61) using the Marquardt's algorithm, the optimum is given by:

$$\underline{\theta}_{opt} = \left[\begin{array}{c} R_s \\ R_r \\ L_m \\ L_f \end{array} \right] = \left[\begin{array}{c} 9.507\ \Omega \\ 4.010\ \Omega \\ 0.4364\ \text{H} \\ 0.0751\ \text{H} \end{array} \right]$$

Figure 1.3 shows the estimation $\underline{\widehat{\theta}}$ according to the iterations of the identification algorithm.

At the optimum, an estimation of the noise variance is obtained with:

$$\widehat{\sigma}^2_b = \frac{J_{opt}}{2\,(K - N)} = 0.0462$$

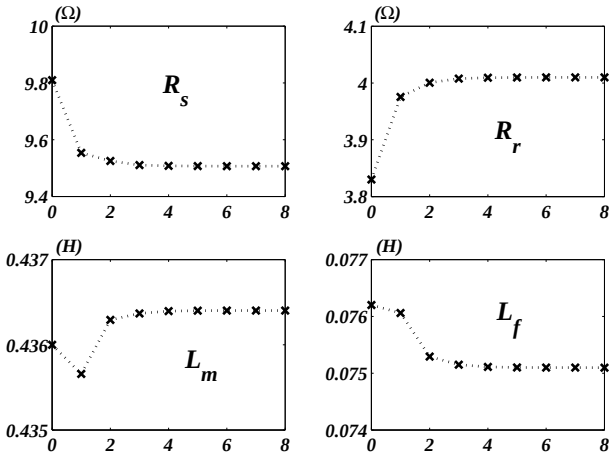

Figure 1.3. *Estimation of $\widehat{\underline{\theta}}$ according to algorithm iterations*

1.4.6. *Output-Error identification and a priori information*

We propose to estimate the same parameters, but with *a priori* information, constituted by the average of ten preliminary estimations (corresponding to the knowledge of the "healthy" functioning of the machine). For that, the composite criterion is minimized:

$$J_c = (\hat{\underline{\theta}} - \underline{\theta}_0)^T M_0^{-1} (\hat{\underline{\theta}} - \underline{\theta}_0) + \frac{\sum^{K} (i^*_{d_{s_k}} - \hat{i}_{d_{s_k}})^2 + (i^*_{q_{s_k}} - \hat{i}_{q_{s_k}})^2}{\hat{\delta}^2} \quad (1.62)$$

with

$$\underline{\theta}_0 = \begin{bmatrix} R_{s_0} \\ R_{r_0} \\ L_{m_0} \\ L_{f_0} \end{bmatrix} = \begin{bmatrix} 9.81\ \Omega \\ 3.83\ \Omega \\ 0.436\ \text{H} \\ 0.0762\ \text{H} \end{bmatrix}$$

and

$$M_0 = \begin{bmatrix} \sigma^2_{R_S} & 0 & 0 & 0 \\ 0 & \sigma^2_{R_r} & 0 & 0 \\ 0 & 0 & \sigma^2_{L_m} & 0 \\ 0 & 0 & 0 & \sigma^2_{L_f} \end{bmatrix}$$

$$= \begin{bmatrix} 2\ 10^{-3} & 0 & 0 & 0 \\ 0 & 2\ 10^{-4} & 0 & 0 \\ 0 & 0 & 6\ 10^{-7} & 0 \\ 0 & 0 & 0 & 10^{-7} \end{bmatrix}$$

The same data as previously are used; then, we choose:

$$\delta^2 = \widehat{\sigma}^2_b = \frac{J_{opt}}{2\,(K-N)} = 0.0462$$

Minimizing J_C using the Marquardt's algorithm, we obtain $\underline{\theta}_C$:

$$\underline{\theta}_C = \begin{bmatrix} R_s \\ R_r \\ L_m \\ L_f \end{bmatrix} = \begin{bmatrix} 9.667\ \Omega \\ 3.920\ \Omega \\ 0.4366\ \mathrm{H} \\ 0.0762\ \mathrm{H} \end{bmatrix}$$

Figure 1.4 shows the evolution of $\widehat{\underline{\theta}}_C$ according to algorithm iterations: it is obvious that the addition of *a priori* information has significantly accelerated the convergence of the algorithm.

Considering the variance of the *a priori* information (better precision on inductances than on resistances), only resistances estimations are slightly different from those corresponding to $\underline{\theta}_0$.

REMARK. The second part of the criterion enables us to estimate the variance of the noise for $\widehat{\underline{\theta}} = \underline{\theta}_C$. Thus, we obtain $\widehat{\sigma}^2_b = 0.0483$ which is close to the value initially chosen for δ^2: it is thus useless in this case to reiterate the algorithm to determine the optimal value of δ^2.

Nevertheless, the question of the choice of δ^2 must be evoked if initial information on the residuals variance is erroneous. Again with the same data, J_C was initialized with an erroneous value of δ^2 and the estimation of $\underline{\theta}_{opt}$ was iterated. The results are shown on figures 1.5 and 1.6. It can be seen that this iterative process converges almost in only one iteration to $\delta^2 = 0.0483$, whatever the initial value of δ^2.

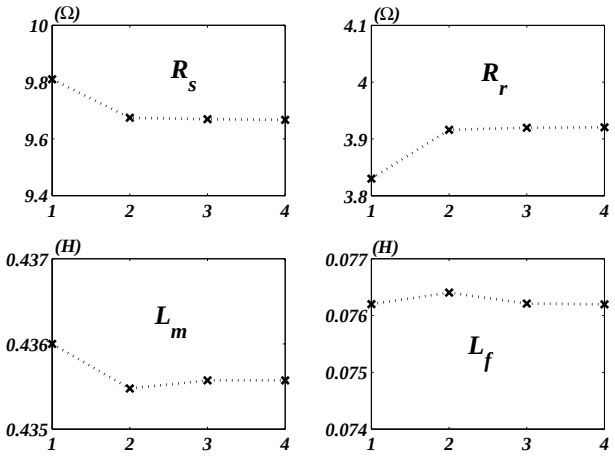

Figure 1.4. *Evolution of $\widehat{\underline{\theta}}_C$ according to algorithm iterations*

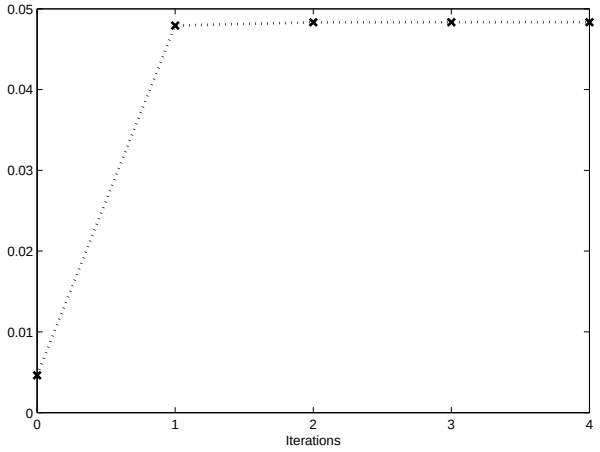

Figure 1.5. *Evolution of δ^2 according to algorithm iterations*

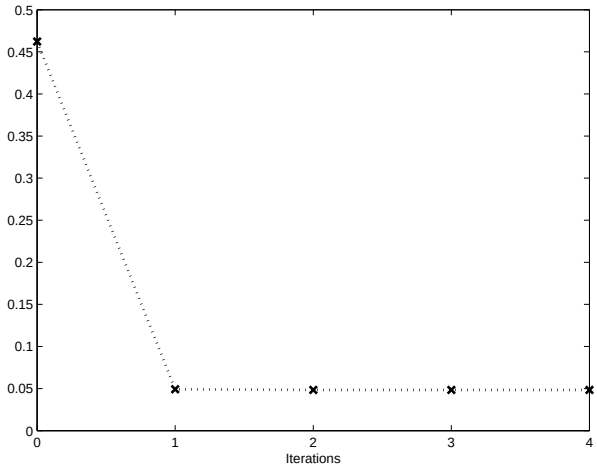

Figure 1.6. *Evolution of δ^2 according to algorithm iterations*

1.5. Fault detection and localization based on parameter estimation

1.5.1. *Introduction*

The fundamental assumption to monitor (or supervise) a system by parameter estimation is that a fault results in the variation of one (or several) characteristic parameter(s) of the system, thus constituting the signature of this fault. According to this assumption, supervising a system involves monitoring of its parameters using an identification algorithm, either off-line (or by parts of samples) or in a recursive way.

In fact, this assumption can easily be invalidated by the fact that this methodology is not able to distinguish a normal parametric variation (possibly foreseeable) from that corresponding to a fault occurring randomly. That is due to the fact that in order to estimate parameters, a model should initially be defined: the first reflex is indeed to use the model of normal operation of the system. However, a fault tends to modify this model and modifies also, in some cases, its structure and in most cases a modeling error is introduced.

Thus, we will propose a methodology again based on parameter estimation, but that combines two characteristics:

– the general model will include a model of safe functioning (or nominal model) and a fault model (specific to each considered fault),

– parameter estimation will be used with *a priori* information, which corresponds to the expertise (or knowledge) of the user on the safe functioning of the system.

1.5.2. *Principle of the method*

The principle of the method is exposed here in the case of linear systems governed by differential equations with constant parameters, although this methodology is general.

Let $H_n(s)$ be a system of nominal transfer function, characterized by a vector $\underline{\theta}_n$. When a fault occurs, a modeling error $\delta H_i(s)$ signing the fault also appears ($\delta H_i(s)$ is characterized by a vector $\underline{\theta}_i$). Thus, the input/output transfer function becomes:

$$H(s) = H_n(s) + \Delta H_i(s) \tag{1.63}$$

The general model of the system, in a fault situation, is shown on figure 1.7, where $b(t)$ is a random perturbation, $u(t)$ is the input and $y^*(t)$ is the measured output.

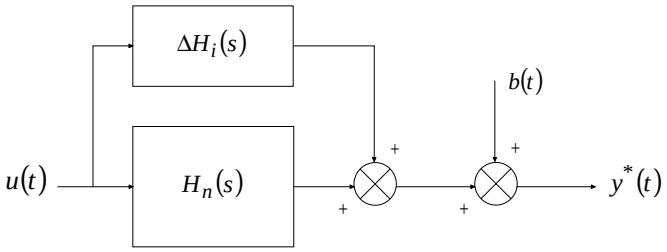

Figure 1.7. *General model of the system corresponding to the fault d_i*

The nominal model $H_n(s)$, or safe functioning model, summarizes the user expertise on the functioning of the system, *i.e.* the knowledge on the nominal parameters $\widehat{\underline{\theta}}_n$ and on their variance $Var\{\widehat{\underline{\theta}}_n\}$, as well as noises affecting the output, *i.e.* their variance σ_b^2. In addition, the modeling error $\delta H_i(s)$ must constitute a true signature of the fault, not only by its structure but as well as its parameters $\underline{\theta}_i$.

The general model of the system $H(s)$ is then composed of a term of "common mode" (the nominal model $H_n(s)$) and of a term of "differential mode" (the fault model $\delta H_i(s)$) only sensitivized when a fault d_i appears. In addition, the nominal model must take into account foreseeable variations of the parameters whereas the fault model must for its part remain insensitive to these same variations.

Finally, the nominal model must include the expertise of the user, *i.e.* summarized by $\{\widehat{\underline{\theta}}_n, var\{\widehat{\underline{\theta}}_n\}\}$. Thus, this methodology is naturally linked to identification with *a priori* information. An extended parameter vector is thus defined:

$$\underline{\theta}_e = \begin{bmatrix} \underline{\theta}_n \\ \underline{\theta}_i \end{bmatrix} \tag{1.64}$$

Moreover, an extended covariance matrix is defined:

$$Var\{\underline{\theta}_e\} = \begin{bmatrix} Var\{\underline{\theta}_n\} & 0 \\ 0 & Var\{\underline{\theta}_i\} \end{bmatrix} \tag{1.65}$$

The *a priori* knowledge can be essentially defined on the nominal model. Then, we obtain:

$$\underline{\theta}_{e_0} = \begin{bmatrix} \hat{\underline{\theta}}_n \\ 0 \end{bmatrix} \tag{1.66}$$

and

$$Var\left\{\underline{\theta}_{e_0}\right\} = \begin{bmatrix} \sigma^2_{\theta_{1n}} & & & & & & \\ & \ddots & & & 0 & & \\ & & \sigma^2_{\theta_{Nn}} & & & & \\ & & & \infty & & & \\ & 0 & & & \ddots & & \\ & & & & & \infty & \end{bmatrix} \tag{1.67}$$

Notice that $Var\{\widehat{\underline{\theta}}_{n_0}\}$ takes into account only the diagonal terms resulting from $Var\{\widehat{\underline{\theta}}_n\}$. In addition, the terms $\sigma^2_{\theta_{jn}}$, resulting from a safe functioning, must be overestimated in order to tolerate foreseeable parameter variations (for example according to the temperature or to the magnetic state).

On the other hand, as one does not know if the fault will occur, its *a priori* value θ_i is null while its initial variance is infinite (or very large). Thus, the optimization algorithm responsible for the minimization of the criterion:

$$J_C = \left(\hat{\underline{\theta}}_e - \underline{\theta}_{e_0}\right)^T Var\left\{\underline{\theta}_{e_0}\right\}^{-1} \left(\hat{\underline{\theta}}_e - \underline{\theta}_{e_0}\right) + \frac{1}{\sigma_b^2} \sum_{k=1}^{K} \left(y_k^* - \hat{y}_k\left(\hat{\underline{\theta}}_e\right)\right)^2 \tag{1.68}$$

will affect the expertise of the user in the nominal model, by tolerating foreseeable variations (included in $\sigma^2_{\theta_{jn}}$), and will be on the other hand very sensitive to the variations $\underline{\theta}_i$ of the differential model, characteristic of the fault d_i.

1.5.3. *Simulations*

In order to specify the methodology which has been just presented, let us consider an academic electrical example.

A reel of N turns winded up on a magnetic circuit of section S and of average length l (see figure 1.8) is considered.

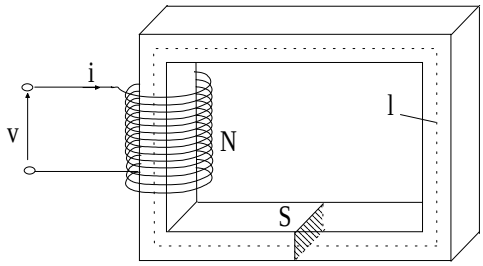

Figure 1.8. *Magnetic circuit*

The iron of the magnetic circuit is supposed to be characterized by the relation $B = \mu H$ (where $\mu = cst$). Neither the iron saturation nor the magnetic hysteresis is taken into account.

In addition, it is assumed that the iron losses are negligible (at first approximation). Then, according to the Ampere's theorem: $H = \frac{Ni}{l}$, the total flux ϕ is given by:

$$\phi = NBS = \mu \frac{N^2 S}{l} i \tag{1.69}$$

The reel inductance L can be defined according to $\phi = Li$:

$$L = \mu \frac{N^2 S}{l} \tag{1.70}$$

Moreover, the reel resistance R is proportional to the length of the electric wire, *i.e.* to the number of turns. One can thus write that $R = k_1 N$ and $L = k_2 N^2$.

Let us define a problem (academic) where the fault is constituted by a variation δN of the number of turns of the reel (compared to nominal number N): in practice, it can be due to a winding with commutation of the number of turns.

Then, if N varies by ΔN:

– R varies by ΔR with $\Delta R = k_1 \Delta N$,
– L varies by ΔL with $\Delta L = \left(\frac{dL}{dN}\right) \Delta N$, *i.e.* $\Delta L = 2k_2 N \Delta N$,

thus, the variations of R and L are linked.

In the nominal state of the reel:

$$\frac{L}{R} = \tau = \frac{k_2 N}{k_1} \tag{1.71}$$

where τ is the time constant of the reel while:

$$\frac{\Delta L}{\Delta R} = \frac{2k_2 N}{k_1} = 2\tau \tag{1.72}$$

Then, a nominal model (nominal impedance) of the reel can be define by:

$$Z_n(s) = R_n + L_n s \tag{1.73}$$

and a fault model by:

$$\Delta Z(s) = \Delta R + \Delta L(s) = \Delta R(1 + 2\tau s) \tag{1.74}$$

which lead to the extended model of the reel:

$$Z(s) = Z_n(s) + \Delta Z(s) = R_n + L_n s + \Delta R(1 + 2\tau s) \tag{1.75}$$

In addition, R and L can vary without the appearance of a fault, for example with changes in the temperature or in the magnetic state of iron. Since μ was assumed constant, we cannot consider variation of L without modification of this assumption. On the other hand, we can consider a variation of the resistance (alone) under the effect of heating, therefore of an increase in temperature T. Then

$$R(T) = R_n + \Delta R(T) \tag{1.76}$$

and

$$L(T) = L_n \tag{1.77}$$

REMARK. An important problem concerns the identifiability of the parameters of the fault model. For this, let us consider the sensitivity functions. Knowing that $I(s) = \frac{U(s)}{Z(s)}$, it can easily be shown that:

$$\begin{cases} \sigma_{R_n}(s) = L\{\sigma_{R_n}(t)\} = \frac{-I(s)}{Z(s)} \\ \sigma_{L_n}(s) = L\{\sigma_{L_n}(t)\} = \frac{-s\,I(s)}{Z(s)} \\ \sigma_{\Delta R}(s) = L\{\sigma_{\Delta R}(t)\} = \frac{-(1+2\tau s)I(s)}{Z(s)} \end{cases} \tag{1.78}$$

where $L\{.\}$ is the Laplace's transform.

Then:

$$\sigma_{\Delta R}(t) = \sigma_{R_n}(t) + 2\tau\ \sigma_{L_n}(t) \tag{1.79}$$

Since $\sigma_{\Delta R}(t)$ is a linear combination of $\sigma_{R_n}(t)$ and $\sigma_{L_n}(t)$, the fault parameter ΔR is unidentifiable as the pseudo-hessian $J''_{\theta\theta} \approx 2\sum_{k=1}^{K} \underline{\sigma}_{k,\underline{\theta}_i} \underline{\sigma}^T_{k,\underline{\theta}_i}$ of the direct method is non invertible.

On the other hand, when the composite criterion J_C is used, which incorporates *a priori* information $\{\underline{\theta}_0, M_0\}$, we obtain the corresponding hessian [1.43]:

$$J''_{C_{\theta\theta}} \approx 2M_0^{-1} + \frac{J''_{\theta\theta}}{\sigma_b^2}$$

which is now invertible thanks to M_0 and the fault parameter ΔR becomes identifiable.

This example shows clearly the interest to associate a model dedicated to a type of fault with the knowledge obtained on the safe functioning in the framework of a strategy of fault detection.

1.5.4. *Numerical simulations*

1.5.4.1. *Study of the safe functioning*

Let us consider a reel characterized by $R = 4\,\Omega$ and $L = 0.1\,\text{H}$. The functioning of this reel was simulated numerically with $T_e = 1$ ms and a PRBS input voltage. The current output was disturbed by a white noise in such a way that the signal to noise ratio $S/B = 10$. An input/output data file is then constituted (see figure 1.9).

Using the Output-Error identification algorithm (without *a priori* knowledge), the values of R and L have been estimated. These values will be the basis of our expertise on the safe functioning.

Thus, for the model $Z_n(s) = R_n + L_n\, s$, we have:

$$\begin{cases} \hat{R}_n = 4.012\ \Omega & \sigma_{R_n} = 3.85\ 10^{-2}\ \Omega \\ \hat{L}_n = 0.0989\ \text{H} & \sigma_{L_n} = 1.81\ 10^{-3}\ \text{H} \\ \hat{\tau}_n = 0.0247\ \text{s} & \hat{\sigma}_b^2 = 1.64\ 10^{-3} \end{cases}$$

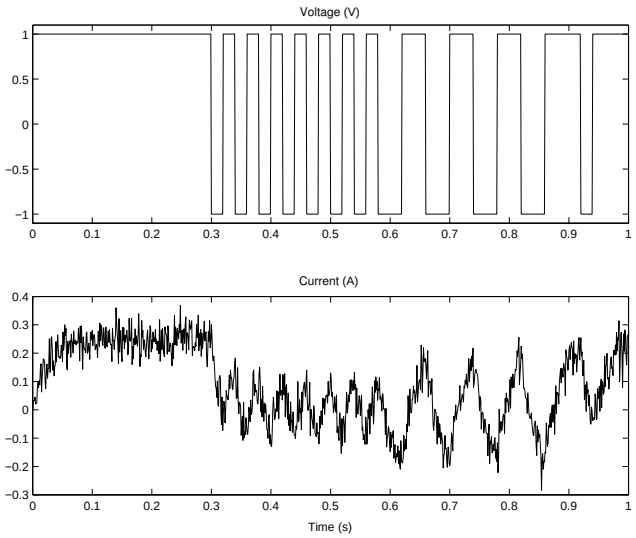

Figure 1.9. *Input/output data*

The prior information is then defined by:

$$\begin{cases} R_0 = 4.012\ \Omega, \quad L_0 = 0.0989\ \text{H}, \quad \tau = 0.0247\ \text{s} \\ \sigma_{R_0} = 1\ \Omega > \sigma_{R_n} \text{ (which authorizes the variations of } R \text{ with the temperature)} \\ \sigma_{L_0} = \sigma_{L_n} \\ \sigma_b^2 = \hat{\sigma}_b^2 \end{cases}$$

1.5.4.2. *Study of the functioning with a fault*

Let us consider a variation of ΔN of the total number of turns, which corresponds to the fault model:

$$Z(s) = R_n + L_n\, s + \Delta R\,(1 + 2\tau\, s) \tag{1.80}$$

with $\underline{\theta}_e^T = \left[\ R_n \quad L_n \quad \Delta R\ \right]$.

Let us consider in addition variations of R_n due to the temperature, that is to say $\delta R(T)$.

Test	1	2	3	4	5
$R\,(\Omega)$	4	5	5	5	5
$\Delta R\,(\Omega)$	1	0	0.2	1	-0.2
$\widehat{R}\,(\Omega)$	3.966	4.965	4.962	5.066	4.979
$\widehat{L}\,(H)$	0.0990	0.0988	0.0989	0.0989	0.0989
$\Delta\widehat{R}\,(\Omega)$	1.066	0.047	0.172	0.886	-0.148

Table 1.1. *Results of parameter estimation*

For each considered situation, the same input as before is used, but the realization of the white noise was different (however with the same signal to noise ratio). All the results are presented in table 1.1.

Test 1 corresponds to an increase of N (corresponding to $\Delta R = 1\,\Omega$) without temperature variation: the increase of R in the fault model is perfectly detected (taking into account of the noise level).

Reciprocally, test 2 corresponds to a temperature variation (increase of R of the common mode), without variation of δR of the differential mode: only the resistance of the common mode varied.

Tests 3, 4 and 5 correspond to simultaneous variations of the temperature and of the number of turns (increase or decrease in δR): the results show the corresponding changes of resistances of the common and differential modes, in close connection with their cause (always taking into account of the noise level) and independently of their amplitude.

In conclusion, the association of a fault model (with common and differential modes) and an algorithm of parameter estimation with *a priori* knowledge constitutes a tool for the fault detection, making it possible moreover to effectively distinguish them from the parameter variations of common mode.

1.6. Conclusion

This chapter was devoted to Output-Error identification and particularly to the estimation of physical parameters within the framework of electrical engineering. Two approaches held our attention:

– the traditional approach is the extension of the least squares method to the non-linear systems; despite its computational load, its essential interest lies in its natural immunity against random disturbances thus guaranteeing an unbiased estimator;

– the Bayesian approach makes it possible to include *a priori* knowledge available on the system, mainly when the user has to deal with physical problems; however, this initial knowledge must be completed by the variance information so as to avoid the risk of biasing the estimator.

The Error-Equation approaches, afflicted with a bias inherent to the construction of the regressor, should not be systematically rejected. They enable, despite this bias, initialization of the research of the optimization algorithm (thus avoiding possible secondary optima) and, if required, they can take part in the development of the *a priori* information within the framework of the Bayesian approach.

The two Output-Error techniques of parameter estimation have been tested and compared in the case of the induced machine. In addition, we have proposed a new methodology of faults detection, based on the Bayesian approach (the *a priori* knowledge corresponds to the expertise of the user on the safe functioning of its system) and on the use of a fault model, true signature of this fault. This methodology, validated by a numerical simulation, will be taken again and generalized with the case of the asynchronous machine in the next chapter which is devoted to the detection of its stator and rotor faults.

1.7. Bibliography

[ALM 95] ALMIAH H., Modélisation et identification en ligne des paramètres d'une machine asynchrone saturée en régime statique, Thèse de doctorat, Université de Poitiers, 1995.

[CAR 95] CARON J. P., HAUTIER J. P., *Modélisation et commande de la machine asynchrone*, Editions Technip, 1995.

[CHA 87] CHATELAIN J., *Machine électriques*, Dunod, 1987.

[EYK 74] EYKHOFF P., *System identification, parameter and state estimation*, Wiley, London, 1974.

[FON 95] FONTEIX C., BICKING F., PERRIN E., MARC I., "Haploïd and diploïd algorithms, a new approach for global optimization : compound performances", *Int. J. Systems Sci.*, vol. 26, num. 10, p. 1919-1933, 1995.

[GOO 77] GOODWIN G., PAYNE R., "Dynamic system identification : experiment design and data analysis", *Math. In Sc and Engineering*, vol. 136, Academic Press, 1977.

[GRE 97] GRELLET G., CLERC G., *Actionneurs électriques. Principes, modèles et commande*, Eyrolles, Paris, 1997.

[GRO 00] GROSPEAUD O., Contribution à l'identification en boucle fermée par erreur de sortie, Thèse de doctorat, Université de Poitiers, 2000.

[HIM 72] HIMMELBLAU D. M., *Applied non linear programming*, Mc Graw Hill, 1972.

[ISE 92] ISERMAN R., LACHMANN K., MATKO D., *Adaptative control systems*, Prentice Hall, 1992.

[JEM 97] JEMNI A., Estimation paramétrique des systèmes à représentation continue - Application au génie électrique, Thèse de doctorat, Université de Poitiers, 1997.

[JOH 97] JOHANSEN T., "On Tikhonov reguralization, bias and variance in non-linear system identification", *Automatica*, vol. 33, num. 3, p. 441-446, 1997.

[KAB 97] KABBAJ H., Identification d'un modèle type circuit prenant en compte les effets de fréquence dans une machine asynchrone à cage d'écureuil, Thèse de doctorat, INP de Toulouse, 1997.

[KNU 94] KNUDSEN M., "A sensitivity approach for estimation of physical parameters", 10^{th} *IFAC Symposium on System Identification, SYSID'94*, vol. 2, p. 231-237, 1994.

[LAN 97] LANDAU I. D., KARIMI A., "Recursive algorithms for identification in closed loop: a unified approach and evaluation", *Automatica*, vol. 33, p. 1499-1523, 1997.

[LJU 87] LJUNG L., *System identification: Theory for the user*, Prentice Hall, USA, 1987.

[MAR 63] MARQUARDT D. W., "An Algorithm for least-squares estimation of non-linear parameters", *Soc. Indust. Appl. Math*, vol. 11, num. 2, p. 431-441, 1963.

[MEN 99] MENSLER M., Analyse et étude comparative de méthodes d'identification des systèmes à représentation continue. Développement d'une boîte à outil logicielle, Thèse de doctorat, Université de Nancy I, 1999.

[MOR 99] MOREAU S., Contribution à la modélisation et à l'estimation paramétrique des machines électriques à courant alternatif: Application au diagnostic, Thèse de doctorat, Université de Poitiers, 1999.

[NOU 89] NOUGIER J., *Méthodes de calcul numérique*, Masson, Paris, 1989.

[PET 81] PETERKA V., "*Trends and Progress in System Identification*", Chapter Bayesian approach to system identification, p. 239-304, Editions P. Eykhoff, Pergamon, Oxford, 1981.

[RAD 84] RADIX J., *Filtrage et lissage statistiques optimaux linéaires*, Cepadues, Toulouse, 1984.

[RIC 71] RICHALET J., RAULT A., POULIQUEN R., *Identification des processus par la méthode du modèle*, Gordon and Breach, 1971.

[RIC 91] RICHALET J., *Pratique de l'identification*, Hermès, Paris, 1991.

[SAY 98] SAYED A., KAILATH T., "*Digital Signal Processing Handbook*", Chapter Recursive Least-Squares adaptive filters, CRC Press, 1998.

[SöD 83] SÖDERSTRÖM T., STOICA P., *Instrumental variable methods for system identification*, Springer Verlag, Berlin, 1983.

[TIK 77] TIKHONOV A., ARSENIN U., *Solutions of ill-posed problems*, Winston, Washington, D.C., 1977.

[TRI 88] TRIGEASSOU J. C., *Recherche de modèles expérimentaux assistée par ordinateur*, Technique et Documentation Lavoisier, Paris, 1988.

[TRI 01] TRIGEASSOU J. C., POINOT T., "*Identification des systèmes*", Chapter Identification des systèmes à représentation continue - Application à l'estimation de paramètres

physiques, p. 177-211, Traité Information, Commande, Communication - Section Systèmes Automatisés, sous la direction de I.D. Landau et A. Besançon-Voda, Hermès, Paris, 2001.

[TUL 93] TULLEKEN H., "Grey-box modeling and identification using physical knowledge and Bayesian techniques", *Automatica*, vol. 29, num. 2, p. 285-308, 1993.

[WAL 97] WALTER E., PRONZATO L., *Identification of parametric models from experimental data.Communications and Control Engineering Series*, Editions Springer, 1997.

[YOU 70] YOUNG P., "An instrumental variable method for real-time identification of noisy process", *Automatica*, vol. 6, p. 271-287, 1970.

Index